

Faculty of Science

University of Split

# **Physical virology**

Mentor: prof. dr.sc. Rudolf Podgornik

Student : Martina Požar

## TABLE OF CONTENTS

Physical virology.....	1
TABLE OF CONTENTS.....	2
1. INTRODUCTION .....	3
2. THE GEOMETRY OF A VIRAL CAPSID .....	4
3. VIRAL ASSEMBLY .....	6
3.1 Capsid proteins.....	6
3.2 Subunit interactions .....	7
3.3 Thermodynamic theory of capsid assembly .....	8
4. ELECTROSTATICS OF VIRUSES.....	11
4.1 Electrostatics of ssRNA viruses .....	14
5. MECHANICAL PROPERTIES OF VIRUSES.....	15
LITERATURE.....	18
PICTURES.....	18

# 1. INTRODUCTION

The word virus comes from the Latin word *virus* which means toxin or poison. That's very telling, since a virus is a microscopic organism which is connected with various diseases that impact the lives and health of humans. Viruses also affect animals, plants and bacteria. It's easy to see why it's important to understand viruses and their formation processes - that knowledge would become a means of developing antiviral therapies that can block virus infections, or redesigning viruses as drug delivery vehicles that can assemble around their cargo and disassemble to deliver it without requiring explicit external control.

An important characteristic of viruses is their inability to reproduce without a host cell. In that regard, viruses are considered parasites, because they contact a host cell with the design to insert their genetic material into the host and take over the host's functions. The infected cell continues to reproduce, but it reproduces more viral genetic material and protein instead of its usual products.

In general, a virus consists of at least two components: the genome (DNA or RNA; single-stranded or double stranded) and a capsid - a protein shell that protects the genome from outer influences. There exists a wide array of viruses, which range from simple (such as the Tobacco mosaic virus (TMV), depicted in Fig. 1) to complex (every virus whose morphology consists of complex combinations of structures, for example the T4 bacteriophage, also depicted in Fig. 1).

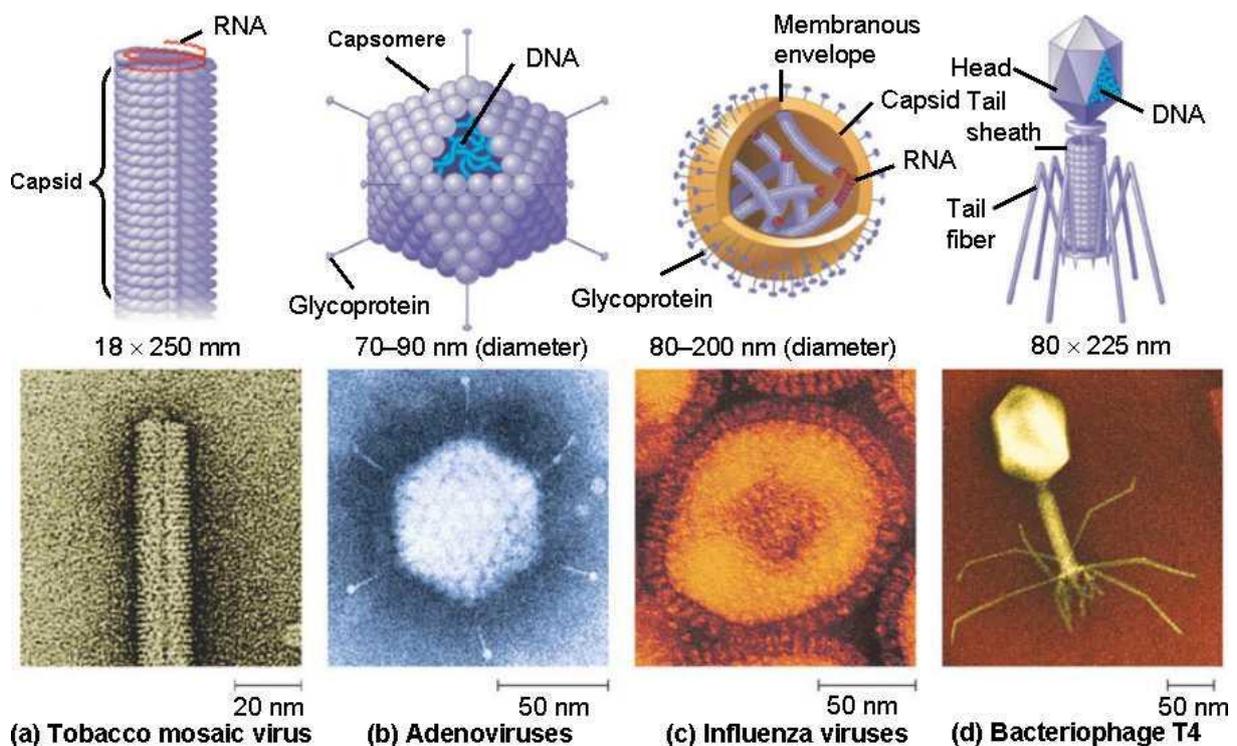


Figure 1 "Anatomies" of different viruses: a) TMV, b) Adenoviruses, c) Influenza viruses, d) T4 bacteriophage

This report shall explore the structure of viral capsids, the self assembly of viruses, with accent on the energies governing the process, and the mechanical properties of both an assembled capsid and a virus in general.

## 2. THE GEOMETRY OF A VIRAL CAPSID

Since the viral genome is supposed to be enclosed in a protective protein shell, that requirement constrains the length of the genome. The resulting consequence is the limited number of protein sequences it can encode. The notable scientists Watson and Crick (of the structure of DNA fame) first made the hypothesis that virus capsids are made of numerous copies of one or a few protein sequences, which are usually arranged with a high degree of symmetry in the assembled capsid.

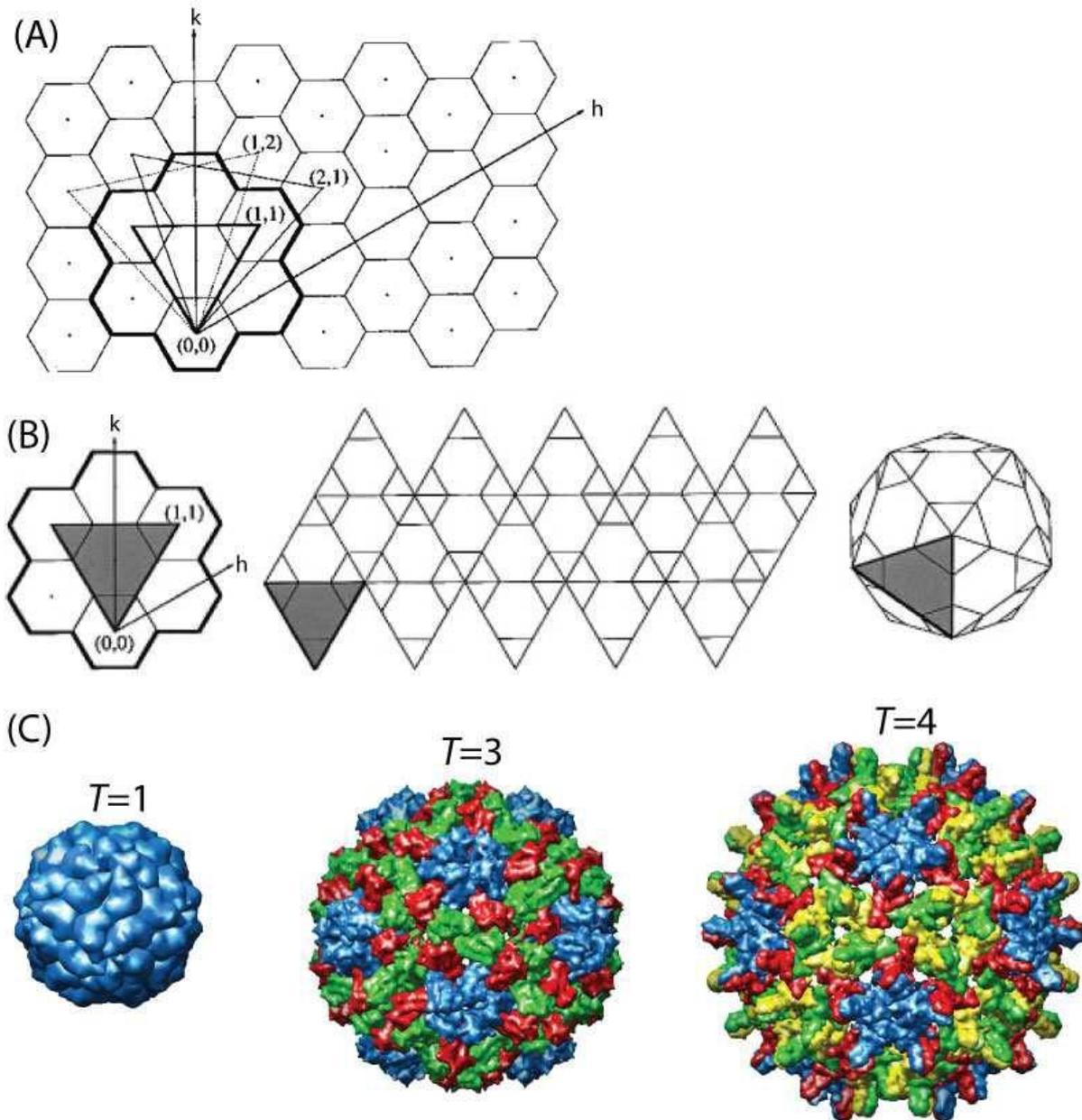
Most viruses can be classified as either rodlike or spherical. Rodlike viruses (for example the aforementioned TMV) have capsids arranged with helical symmetry around the nucleic acid. The number of protein copies comprising a helical capsid is arbitrary and therefore a helical capsid can accommodate a nucleic acid of any length. The other group, spherical viruses, have capsids arranged with icosahedral symmetry. Icosahedral capsids are limited by the geometric constraint that at most 60 identical subunits can be arranged into a regular polyhedron. Thus the assembly of capsids in spherical viruses poses an interesting problem.

Early structural experiments indicated that many spherical capsids contain multiples of 60 proteins. One of the propositions on how multiples of 60 proteins can be arranged with icosahedral symmetry was given by Caspar and Klug. They made geometrical arguments to support their hypothesis. They argued that protein subunits can be grouped into so called capsomeres - morphological units usually comprised as pentameres and hexameres. Icosahedral symmetry requires exactly 12 pentameres, located at the vertices of an icosahedron inscribed within the capsid. A complete capsid is comprised of  $60T$  subunits, where  $T$  is the 'triangulation number', which is equal to the number of distinct subunit conformations. The Caspar Klug (C-K) classification system can be obtained starting from a hexagonal lattice (shown in Figure 2A). The edge of the icosahedral facet is defined by starting at the origin and stepping distances  $h$  and  $k$  along each of the respective lattice vectors. There is an infinite series of such equilateral triangles corresponding to integer values of  $h$  and  $k$ . The area of such a triangle (for unit spacing between lattice points) is given by  $T/4$ , where  $T$  is the triangulation number defined as:

$$T = h^2 + hk + k^2.$$

A structure with icosahedral symmetry is made of 20 identical facets, which are equilateral triangles (shown in Figure 2B). The facets themselves comprise at least 3 identical asymmetric units (asu). The smallest possible triangle, with  $T = 1$  is made of 3 asu, so the total number of asu in the facet is  $3T$ , and the total number of asu in the icosahedron is  $60T$ . It's already said that a CK icosahedral shell consists of 12 pentameres located at equidistant sites on the icosahedral vertices. But, it has a further  $10(T - 1)$  hexameres - with  $T = 1, 3, 4, 7; \dots$  - located in between the pentameres. The individual asu's are not identical for  $T > 1$  since they have different local environments. A good deal of icosahedral viral capsids with  $T > 1$  are comprised of only a single protein copy. That means that the protein must

adopt different configurations depending on its local environment. Caspar and Klug also hypothesized that because the local environments 'quasi-equivalent', the proteins in different environments could interact through the same interfaces. Their hypothesis was proven to be true for many icosahedral viruses, with structural differences between proteins at different quasi-equivalent sites often limited to loops and N- and C-termini. However, there are exceptions to that rule. There can be proteins with extensive conformational changes or even different sequences at different sites.



**Figure 2** The geometry of icosahedral lattices. A) Triangular facets and their construction on a hexagonal lattice. B) Construction of a  $T=3$  lattice. C) Icosahedral capsids. They belong to: satellite TMV ( $T=1$ ), cowpea chlorotic mottle virus ( $T=3$ ), human hepatitis B virus ( $T=4$ ).

It's interesting to note that there exist viruses with non-spherical capsids that display elements of icosahedral symmetry. One such is the mature HIV virus - the HIV capsid has a tubular or conical shape.

### 3. VIRAL ASSEMBLY

Viral assembly is a process that includes: formation of the viral capsid, encapsulation of the nucleic acid in the capsid, acquisition of membrane coats (if the virus has any) and eventual maturation processes. Viral capsids can spontaneously form. That was proven by the experiment Fraenkel-Conrat and Williams had done in 1955. They had demonstrated that a functional TMV virus could be created out of purified RNA and a protein coat, *in vitro*.

There is a big difference between the pathways of nucleic acid encapsulation between viruses with single-stranded or double-stranded genomes. Viruses with single-stranded genomes (the best studied of which have ssRNA genomes) usually assemble spontaneously around their nucleic acid in a single step. Such viruses are small spherical plant viruses, like satellite TMV, the bacteriophage MS2, and animal viruses such as nodavirus. Frequently, the RNA is required for assembly at physiological conditions, whereas the capsid proteins can assemble without RNA into empty shells *in vitro* under different or pH.

Double-stranded genomes pose a greater challenge. Since a double-stranded genome is stiff (the persistence length of dsDNA is 50 nm) and has a high charge density, it requires a two-step process to encapsulate it within a capsid. Basically, an empty protein shell is assembled first, followed by packaging via ATP hydrolysis and/or complexation with nucleic acid folding proteins. In that category, the most studied viruses are dsDNA viruses, such as the tailed bacteriophages, herpes virus and adenovirus. These viruses assemble an empty capsid, without requiring a nucleic acid at physiological conditions, and a molecular motor which inserts into one vertex of the capsid. The molecular motor uses the hydrolysis of ATP to pump the DNA into the capsid.

The assembly of virus capsids from the coat proteins is a thermodynamic process for a vast majority of viruses. Together with the aforementioned TMV, HBV, Human Papilloma virus (HPV), Cowpea Chlorotic Mottle virus, Brome Mosaic virus, Broad Bean Mottle virus and Sindbis virus are all examples of viruses where the coat proteins spontaneously form capsids in aqueous solution under the right conditions of concentration, salinity, pH, and temperature. But first, it's necessary to analyze not only the formation process of an empty capsid, but also the proteins that build a viral capsid, and the interactions between the subunits of a capsid.

#### 3.1 Capsid proteins

Viral DNA or RNA molecules are typically negatively charged. In order to compactly and efficiently pack negatively charged genome in to a capsid, the capsid proteins in contact with the genome are oftentimes positively charged. The opposite charges of the genome and capsid proteins increase the electrostatic interaction of the complex and decrease its energy, which leads to easier assembly. That explains why viral genome codes for positively charged proteins.

Virus capsid proteins seem to have a few characteristic structural motifs. Most small, non-enveloped viruses share common protein fold - an eight-strand antiparallel  $\beta$ -barrel (so called „jelly roll fold“). That motif can be found in RNA picornaviruses, DNA parvoviruses, unrelated families of non-

eveloped RNA plant viruses, DNA polyomaviruses and papillomaviruses, and some DNA bacteriophages. But not all non-enveloped spherical viruses have in common this  $\beta$ -barrel subunit fold. The small RNA bacteriophage MS2 has a five-strand  $\beta$ -sheet flanked by two C-terminal  $\alpha$ -helices. (This fold is characteristic for other members of the Leviviridae family such as Q $\beta$ , R17, and PP7.) Many large multi-component phages share a common fold that was first discovered in bacteriophage HK97. In the Hepatitis B virus (HBV), an enveloped DNA virus, the capsid protein has a unique  $\alpha$ -helical fold.

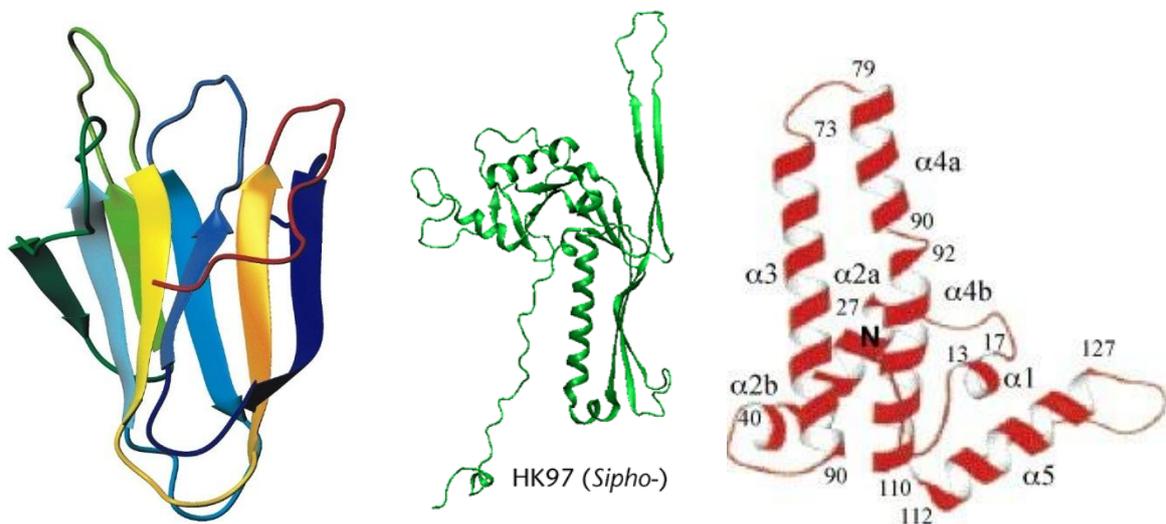


Figure 3 Left - "jelly roll fold". Middle - structural motif of HK97 virus. Right - helix motif of HBV.

Even though icosahedral geometry of a capsid offers stability and structure to a virus, not all viruses share that trait. Retroviruses are an interesting group in that sense, because they have irregular capsids, with a predominantly helical capsid protein.

### 3.2 Subunit interactions

For an icosahedron to be assembled, there has to be hexameric and pentameric contacts. Contact domains must be capable of sustaining interactions in the necessarily different local geometries. The assembly of virus capsids tends to be driven by the burial of hydrophobic surface area at the inter-subunit contact points. That is consistent with the observation that viral assembly is driven by an increase in system entropy. Usually, a single capsid inter-subunit contact buries some 1750  $\text{\AA}^2$  of surface area, which is a relatively small contact area (although bigger than a simple crystal contact). That implies that, while the intermediates in viral assembly are unstable, a network of otherwise relatively weak interactions stabilizes the whole formation.

The HBV capsid formation may be used as an example, because many thermodynamic models of assembly were used for explaining the formation of that virus. HBV forms a T = 4 capsid based on hydrophobic contacts, with each asymmetric unit comprised of two copies of the homodimeric capsid protein (depicted in Fig. 4, the middle). The structure of this capsid has been solved to 3.3  $\text{\AA}$ .

The assembly of HBV is driven by increasing ionic strength, temperature, and capsid protein concentration. On average a single HBV subunit-subunit contact buries  $\sim 1500 \text{ \AA}^2$  (that estimate was obtained by comparing experimental results with the crystal structure).

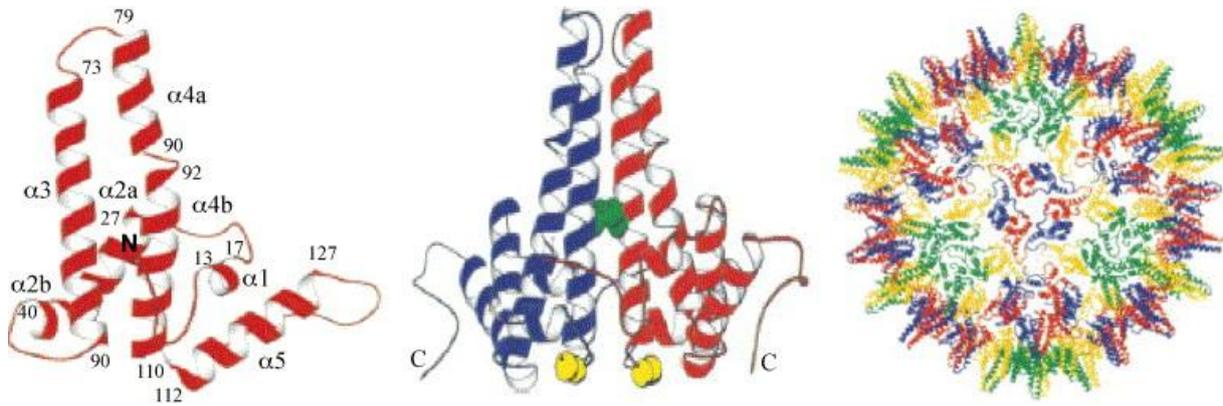


Figure 4 Left - HBV capsid monomer. Middle - HBV capsid dimer. Right - HBV capsid.

### 3.3 Thermodynamic theory of capsid assembly

Plainly speaking, to have a virus self-assemble, states with capsids must be energetically favorable, i.e. they have to have lower free energy than states with only free subunits. Since the disordered units (and RNA or other components in some cases) form an ordered capsid structure, that means that their translational and rotational entropy decreases. Such events, then, must be driven by extremely favorable interactions among subunits and any other components if they can overcome the decrease of entropy.

Protein-protein interactions are very important for the whole assembly process. As discussed in the previous section, capsid proteins are usually highly charged and possess binding interfaces that bury large hydrophobic areas. Together with hydrophobic interaction, the assembly also depends on electrostatic, van der Waals, and hydrogen bonding interactions. (Covalent interactions do not participate in the assembly, but rather in the maturation processes of some viruses, like the HK97 bacteriophage.) All of the interactions mentioned above are short-ranged under assembly conditions. Van der Waals interactions and hydrogen bonds are measured in the scale of few angstroms. Electrostatic interactions are measured on the scale of the Debye length,  $\lambda_D \approx \frac{0,3}{C_{salt}^{1/2}}$ , where  $\lambda_D$  is measured in nanometers and the salt concentration  $C_{salt}$  measured in molar units. The hydrophobic interactions are similarly characterized by a length scale of approximately a 0,5 – 1 nm.

In many cases, hydrophobic interactions primarily drive the assembly, weakened by electrostatic interactions with directional specificity imposed by van der Waals interactions and hydrogen bonding  $\text{\AA}$  length scales. That was proven by copious experiments done on the HBV capsids, which showed that the thermodynamic stability of HBV capsids increases with both temperature and ionic strength. The former led to the conclusion that hydrophobic interactions are the dominant driving force, while the

latter suggests that the salt screens repulsive electrostatic interactions which oppose protein association.

Taking into account the interactions between the subunits, it's possible to calculate the free energy for a system of identical protein subunits assembling to form empty  $T = 1$  capsids. Also, an assumption can be made: there is one dominant intermediate species for each number of subunits  $n$ . The total free energy  $F_{EC}$  for a system of subunits, intermediates, and capsids in solution can be written as:

$$(1) \quad \frac{F_{EC}}{k_B T} = \sum_{n=1}^N k_B T \rho_n [\log(\rho_n v_0) - 1] + \rho_n G_n^{cap},$$

where  $v_0$  is a standard state volume,  $\rho_n$  is the density of intermediates with  $n$  subunits, and  $G_n^{cap}$  the interaction free energy of such an intermediate. A plausible model for the interaction free energy is:

$$(2) \quad G_n^{cap}(g_b) = \sum_{j=1}^n (n_j^c g_b) - T S_n^{degen},$$

where  $n_j^c$  is the number of new subunit-subunit contacts formed by the binding of subunit  $j$  to the intermediate,  $g_b$  is the free energy for such a contact, and  $S_n^{degen}$  accounts for degeneracy in the number of ways subunits can bind to or unbind from an intermediate. These terms are specified by the geometry of the capsid. To obtain the equilibrium concentration of intermediates we minimize  $F_{EC}$  subject to the constraint that the total subunit concentration  $\rho_T$  is conserved:

$$(3) \quad \sum_{n=1}^N n \rho_n = \rho_T.$$

This leads to the law of mass action (LMA) result for intermediate concentrations:

$$(4) \quad \begin{aligned} \rho_n v_0 &= \exp[-\beta(G_n^{cap} - n\mu)] \\ \mu &= k_B T \log(v_0 \rho_1), \end{aligned}$$

with  $\mu$  the chemical potential of free subunits and  $\beta = 1/k_B T$ . Thanks to the constraint (3), (4) must be solved numerically. The result for a model dodecahedral capsid comprised of 12 pentagonal subunits is shown in Fig. 5 for several values of the binding energy  $g_b$ . (In all cases, the capsid protein is in a state of either free subunits or complete capsids. This prediction, which is analogous to the result for spherical micelles with a preferred diameter is generic to any description of an assembling structure in which the interaction free energy  $G_n^{cap}$  is minimized by one intermediate size ( $n = N$ ) and the total subunit concentration is conserved.)

Since intermediate concentrations are negligible at equilibrium, the equations of capsid assembly thermodynamics can be simplified using the two-state approximation, the first state being free subunits and the second being complete capsids. In that case, the total subunit concentration is:

$$(5) \quad \rho_T = \rho_1 + N \rho_N.$$

Defining the fraction of subunits in capsids as  $f_c = \frac{N \rho_N}{\rho_T}$ , and combining equations (5) and (4), it is obtained:

$$(6) \quad \frac{f_c}{1-f_c} = N (v_0 \phi_T)^{N-1} e^{-\beta G_n^{cap}}.$$

(A)	n	Model	Build		$S_n$	c	$K'_n$
			Up	Down			
	1		—	—	—	—	—
	2		5/2†	1	5/2	1	$e^{-1\Delta G_2^*/RT}$
	3		2	3	2/3	2	$e^{-2\Delta G_3^*/RT}$
	4		3	2	3/2	2	$e^{-2\Delta G_4^*/RT}$
-----							
	11		2	5	2/5	4	$e^{-4\Delta G_{11}^*/RT}$
	12		1	12	1/12	5	$e^{-5\Delta G_{12}^*/RT}$

Figure 5 Assembly model of a dodecahedral capsid and the statistical weights associated with symmetries of the intermediates.

If  $N \gg 1$ , (6) yields:

$$\frac{f_c^{1/N}}{1 - f_c} = \frac{\rho_T}{\rho^*}$$

$$(7) \quad \rho^* v_0 = \exp\left(\beta \frac{G_n^{cap}}{N-1}\right) N^{-1/(N-1)} \approx \exp(\beta G_n^{cap}/N),$$

where  $\rho^*$  is the pseudo-critical subunit concentration. In the asymptotic limits, equation (7) reduces to:

$$f_c \approx \left(\frac{\rho_T}{\rho^*}\right)^N \ll 1, \text{ for } \rho_T \ll \rho^* \text{ and}$$

$$(8) \quad f_c \approx 1 - \frac{\rho^*}{\rho_T}, \text{ for } \rho_T \gg \rho^*.$$

The solution for (8) is shown in Figure 6, for three values of capsid size  $N$ . By increasing the total subunit concentration  $\rho_T$  or decreasing  $\rho^*$ , the fraction of subunits in complete capsids  $f_c$  at equilibrium is always increased.

As for viruses with a higher  $T$  number, the equations described above can be extended to describe capsids with larger  $T$  numbers, since icosahedral capsids comprise  $T$  different subunit conformations.

In those cases, the capsid free energy,  $G_n^{\text{cap}}$  must be extended to include conformation energies and contact free energies  $g_b$  which depend on the subunit conformation or species.

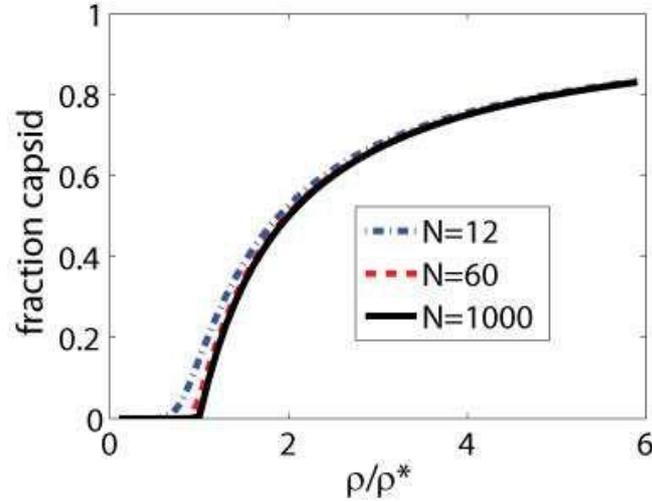


Figure 6 Fraction capsid  $f_c$  as a function of  $\frac{\rho}{\rho^*}$ , as predicted by Eq. 8, for three values of  $N$ .

#### 4. ELECTROSTATICS OF VIRUSES

As discussed in the last chapter, proteins that build a viral capsid are usually positively charged. Thus, the simplest representation of the protein charge distribution would be a positive charge homogeneously distributed on a protein. A homogeneous distribution of positive charge would have to keep proteins apart, and therefore inhibit viral self-assembly. But since it's proven that viruses can self-assemble, it's safe to say that some sort of attractive interaction between the proteins must exist. The source of that interaction is a combination of hydrophobic and van der Waals interactions, although the hydrophobic component dominates.

A virus capsid is represented by a uniformly, positively charged (with surface charge density  $s$ ), permeable, infinitely thin sphere of radius  $R$ . The main goal, in order to understand the cohesive energy of a capsid, is to calculate the electrostatic self-energy. That can be done by using the Poisson-Boltzmann approach, i.e. using the Poisson-Boltzmann (PB) equation, a nonlinear partial differential equation that incorporates detailed information about the biomolecular shape and charge distribution. It can be derived by minimizing the appropriate free energy:

$$(9) \quad F_{PB}[\phi(r), \nabla\phi(r), c^i(r)] = \int f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) d^3r.$$

Here, the ions as ideal gas, which adjust to the external potential and contribute to it via their charge density. The quantity  $f_{PB}$  is free energy density:

$$(10) \quad f_{PB}(\phi(r), \nabla\phi(r), c^i(r)) = -\frac{1}{2} \epsilon_0 \epsilon \nabla\phi(r)^2 + \sum_{i=\pm} e_i c^i(r) \phi(r) + \sum_{i=\pm} k_B T [c^i(r) \ln c^i(r) - c^i(r) - (c_0^i \ln c_0^i - c_0^i)] + e_0 \rho_p(r) \phi(r).$$

The value  $e_0\rho_p(r)$  is the charge density of the capsid,  $k_B$  is the Boltzmann constant,  $T$  is temperature,  $c^i$  are the concentrations of salt ions,  $c_0^i$  their bulk concentrations,  $\epsilon_0\epsilon$  the permittivity of water, and  $e_0$  the electron charge. Minimizing the free energy yields the PB equation:

$$(11) \quad -\epsilon\epsilon_0\nabla^2\phi(r) = \sum_{i=\pm} e_i c_0^i e^{-\beta e_i \phi(r)} + e_0\rho_p(r),$$

where  $\beta^{-1} = k_B T$  and  $e_i$  is the charge of the ions ( $e_i = \pm e_0$ ). When the electrostatic potentials in the solutions are small, and if the system is symmetrical, the PB equation can be linearized, which leads to the Debye-Huckel (DH) equation for potential:

$$(12) \quad -\nabla^2\phi(r) = \frac{\beta}{\epsilon\epsilon_0} (\sum_{i=\pm} e_i^2 c_0^i) \phi(r) + \frac{e_0\rho_p(r)}{\epsilon\epsilon_0} + \dots$$

Since the DH equation is linear, it's possible to solve it multiple ways. One can get the self-energy of a uniformly charged sphere by summing the pair DH interactions over the sphere surface (It's postulated that the interaction between the charges  $Q_1$  and  $Q_2$ , separated by  $r_1 - r_2$  in the solution of monovalent ions (with concentration  $c_0$ ) of (relative) dielectric constant  $\epsilon$  is given by the DH potential of the screened exponential form  $U(r_1 - r_2)$ ):

$$(13) \quad F_{DH} = \frac{\sigma}{2} \int dS_1 \int dS_2 U(\vec{r}_1 - \vec{r}_2) = \frac{1}{2} \frac{\sigma^2}{4\pi\epsilon_0\epsilon} \int dS_1 \int dS_2 \frac{\exp(-\kappa|\vec{r}_1 - \vec{r}_2|)}{|\vec{r}_1 - \vec{r}_2|},$$

where  $dS_1$  and  $dS_2$  are infinitesimal elements of the sphere surface centered around vectors  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . The factor of  $\frac{1}{2}$  is to eliminate double counting of the pair interactions. It has to be taken into account that the radii of viruses are usually of the order of 20 nm, while the DH screening length in the physiological conditions is  $\kappa^{-1} \sim 1$  nm. If  $\kappa R \gg 1$ , that means that the range of integration can be cut off on the scale of  $\kappa^{-1}$ . When  $\kappa R \gg 1$ , instead of defining a spherical cap, the integral defines a disk, and makes the two integrations independent:

$$(14) \quad \lim_{\kappa R \gg 1} F_{DH} = \frac{\sigma^2}{4\pi\epsilon_0\epsilon} \frac{1}{2} \int dS_1 \int_0^\infty dr_2 \int_0^{2\pi} d\phi_2 \exp(-\kappa r_2).$$

The self-energy of the capsid is:

$$(15) \quad \lim_{\kappa R \gg 1} F_{DH} = \frac{\pi\sigma^2 R^2}{\epsilon_0\epsilon\kappa}.$$

So, the self-energy  $F$  is the energy required to bring infinitesimal charges from infinite separations in the solution to the capsid. The quantity can be loosely thought of as the electrostatic contribution to the assembly free energy. For a capsid of about  $\sigma \sim 1$  e<sub>0</sub>/nm<sup>2</sup>, and  $c_0 = 100$  mM, the free energy would be  $F_{DH} \sim 10^4$  k<sub>B</sub>T. It's interesting to note the hydrophobic energy of the area of proteins in a capsid of  $R \sim 20$  nm. A capsid of about 20 nm is typically built of 180 protein subunits, which renders the length of protein contacts in such a capsid about 3000 nm. If the capsid happens to be 2 nm thick, the total area exposed to protein contacts is about 6000 nm<sup>2</sup>. Taking into account that the energy of attractive protein interactions per unit area of the exposed protein surface is typically of the order of 10 mJ/m<sup>2</sup>, the final tally for attractive hydrophobic interaction (begotten by multiplication of the estimated exposed area and energy of attractive interactions) is  $F_{HP} \sim 10^4$  k<sub>B</sub>T. The number is in the same range as the estimated number of free energy above, which can be

explained by the fact that viruses, in addition to being able to assemble, need also to disassemble and deliver their genome molecule to the cell.

However, there are more sophisticated models of viruses. To improve the simplest model, one can examine the whole range of  $\kappa R$  values, valid in particular for lower ionic concentrations when  $\kappa < R$ . One can even derive a general formula that describes the capsid electrostatic free energy for any  $\kappa$  within the range of validity of the DH approximation:

$$(16) \quad F_{DH} = \frac{2\pi\sigma^2 R^2}{\varepsilon_0 \varepsilon \kappa [1 + \coth(\kappa R)]}$$

To further refine the model, one has to assume that the electrostatic potential  $\beta\varepsilon_0\phi$  is not so small to be discarded, and to solve Eq. (9) with that in mind. Moreover, the capsid has to have a finite thickness,  $\delta$ . The distribution of charge across the thickness of the capsid is specific: the positive charges are often concentrated on the capsid interior surface, while, typically negative charges are concentrated on the capsid exterior surface. The capsid was treated as a dielectric shell with relative permittivity  $\varepsilon_0$ , with interior and exterior surfaces which are uniformly charged with surface charge densities  $\sigma_1$  and  $\sigma_2$  (as depicted in Fig. 7).

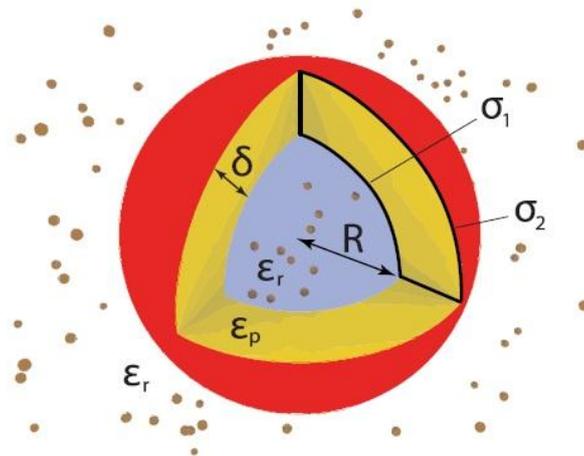


Figure 7 The electrostatic model of a viral capsid, with the thickness  $\delta$ . The brown spheres are salt ions.

In the limit  $\kappa R \gg 1$ , the capsid free energy scales with the second power of capsid radius, both in the DH and PB cases. If we take  $\kappa R \gg 1$ ,  $\varepsilon \geq \varepsilon_p$  and  $\delta \ll R$ , it renders the following equation:

$$(17) \quad F_{DH}(\sigma_1, \sigma_2, \delta) = 2\pi R^2 \frac{\varepsilon_p(\sigma_1 + \sigma_2)^2 + \varepsilon(\sigma_1^2 + \sigma_2^2)\kappa\delta}{\varepsilon_0 \varepsilon \kappa (2\varepsilon_p + \varepsilon\kappa\delta)},$$

which can be simplified further by immersing a virus into physiological conditions ( $\varepsilon \gg \varepsilon_p$  and  $\kappa\delta \sim 1$ ):

$$(18) \quad F_{DH} = \frac{2\pi(\sigma_1^2 + \sigma_2^2)R^2}{\varepsilon_0 \varepsilon \kappa}$$

#### 4.1 Electrostatics of ssRNA viruses

As established previously, viruses with ssRNA often self-assemble. A simple model of an ssRNA virus includes an ssRNA molecule that can be defined as a generic flexible polyelectrolyte. The flexibility of the molecule is characterized by its effective persistence length (the length on which the ssRNA refuses to bend). The Edwards -de Gennes flexible chain model can explain flexible polyelectrolytes. In this case, the effective length is of the order of nucleotide separation ( $a \sim 0.5$  nm). The monomers behave differently than free particles in solutions due to the connectivity of the chain, which also alters the long range interactions between monomers, such as Coulomb interactions in the case of ssRNA. The connectivity is responsible for properties such as adsorption to charged surfaces and the consequent bridging interactions present between two opposed charged surfaces. That's relevant in the case of ssRNA viruses, which encapsulates the ssRNA molecule in a thin shell closely distanced from the interior capsid radius. The set-up of the problem is complex; because one has to deal with interaction energies and chain entropy (the ssRNA chain is flexible, and the elasticity of the chain is purely entropic) to grasp the formation of the adsorbed ssRNA polyelectrolyte layer which forms a shell.

The total free energy in the case of a ssRNA virus has to contain electrostatic interactions due to ions and charges on the polyelectrolyte, and the entropy of the polyelectrolyte chain. The electrostatic part of the free energy,  $F_{es}$ , contains the PB functional given in (10), augmented by the presence of the charges on the polyelectrolyte chain expressed in terms of the polyelectrolyte monomer concentration  $\rho(r)$ :

$$(19) \quad F_{es}[\phi(r), \nabla\phi(r), \rho(r)] = \int f_{es}(\phi(r), \nabla\phi(r), \rho(r)) d^3r - \mu(\int d^3r \rho(r) - N),$$

where  $\mu$  is the Lagrange multiplier enforcing the condition of fixed numbers of monomers,  $N$ , of the polyelectrolyte chain, with

$$(20) \quad f_{es}(\phi(r), \nabla\phi(r), \rho(r)) = f_{PB}(\phi(r), \nabla\phi(r)) - p e_0 \rho(r) \phi(r).$$

In (17),  $f_{PB}(r)$  was defined in (10);  $p e_0$  is the charge per monomer,  $e_0$  is the electron charge and  $0 < p < 1$ . The contribution of entropy of the flexible polyelectrolyte in the free energy expression can be defined with „ground state dominance“ approximation:

$$(21) \quad F_{ent}[\rho(r), \nabla\rho(r)] = k_B T \frac{a^2}{6} \int d^3r \frac{[\nabla\rho(r)]^2}{\rho(r)}.$$

By minimizing the sum of the electrostatic and entropic contributions, one can get to a polyelectrolyte PB equation that can be solved numerically in the spherical geometry of the capsid. However, numerical solutions of the polyelectrolyte PB theory are complicated and won't be touched upon here. The main conclusion one can draw is that an adsorption layer next to the internal positively charged wall of the capsid really does exist. That, in turn, leads to the conclusion that ssRNA should be non-uniformly distributed within the capsid, with a relatively dense surface layer and a depleted core. The existence of the adsorbed layer along the periphery of the capsid actually stabilized the protein shell, because the attractive polyelectrolyte bridging interactions act between different parts of the spherical hypotope.

The question of ssRNA packing inside a capsid is particularly interesting. It has been found that the optimal virus configuration in the physiological regime is such that the total charge on the encapsulated ssRNA is comparable to the charge on the capsid, being equal as the salt concentration decreases. It's a quite logical argument, since the ssRNA can completely screens the protein charges if the charges on the capsid and on the ssRNA molecule are equal. So in that case, the salt ions almost need not to redistribute at all, especially when the ssRNA and the capsid can be brought in close contact. But that's a quite singular case, because in all other cases, there is an effective charge, that the salt ions must screen by rearranging, increasing the total electrostatic energy of the system in this way. The fact that ssRNA is a polyelectrolyte molecule makes things worse, for each nucleotide is supposed to carry a fixed charge, and the charges are connected, which affects the total charge of the ssRNA molecule. The ssRNA configuration inside the capsid can be obtained from the full polyelectrolyte PB theory, that contain also the polyelectrolyte entropy, by using the approximation of two concentric, spherical shells with opposite charge,  $\sigma \equiv \sigma_2 \approx -\sigma_1$ , separated by  $\sim a$ .

But the easiest way to obtain an estimate of the electrostatic complexation free energy ( $\Delta F_C$ ) of a ssRNA virus in the DH approximation can be obtained from (17), using  $\epsilon_p \approx \epsilon$  and  $\kappa\delta \sim 1$  and subtracting the electrostatic self-energy of the capsid, (15), which results in:

$$(22) \quad \Delta F_C \approx F_{DH}(-\sigma, \sigma, a) - F_{DH}(0, \sigma, 0)\pi \approx \frac{\pi\sigma^2 R^2}{\epsilon_0 \epsilon} (fa - \kappa^{-1}),$$

where  $f$  is a numerical factor depending on  $\kappa$  and  $\delta$  (between 4/3 and 2). The result might not be very accurate in high salt concentrations, but it works in general. Using the parameters  $a = 0.5$  nm,  $fa - \kappa^{-1} \sim 0$ , the value for  $\Delta F_C$  is about  $\sim 0$ . Even a more detailed calculation gives a negative  $\Delta F_C$ , which leads to the conclusion that spontaneous encapsidation of ssRNA is a delicate process that may even be suppressed in thermodynamical equilibrium, so that only empty capsids form. In physiological conditions, it was observed that the complexation free energy is negative so that ssRNA encapsidation happens.

## 5. MECHANICAL PROPERTIES OF VIRUSES

After assembling, the viral capsid proteins are often modified in a process called maturation. It was observed that the capsids of certain bacteriophages undergo a whole sequence of conformational changes and chemical reactions that strengthen the capsid. The shell-maturation steps resemble structural phase transitions in crystals.

It's interesting to look into the resiliency of a virus, or an empty capsid after the assembly, in situations when external force is applied to it. After all, capsids need to be both stable and unstable; stable enough to protect their genome in the extracellular environment, but also unstable enough to release their genome molecules into host cells. Various experiments have been conducted on viruses to determine their mechanical properties, and a very interesting one is using atomic force microscopy (AFM) to probe the virus and check for deformations afterwards.

The relation between the applied force and the resulting change in shell diameter is called the force deformation curve (FDC). Depending on whether or not the capsid returns to its original state after the probe force is removed, it's called a reversible or irreversible deformation. The force measured by a nanoindentation probe results from the fact that the probe forces the viral shell away from a state of minimum free energy. For an interpretation of measured FDCs, it's necessary to make an analogy with the „thin-shell theory“ (TST). TST is used to predict the effects of external forces on thin-walled, hollow macroscopic structures, but it can also be applied to a virus, if the capsid is modelled as a thin, spherical shell of uniform thickness and radius  $R$ . If a full virus is discussed, then it must be taken into account that it encloses genome molecules, therefore an internal osmotic pressure  $\Pi$  must be included, which can be as large as  $\sim 50$  atm.

If the force probe has left an indentation profile  $\zeta(\mathbf{r})$  (defined as the radial inward displacement of the surface of the sphere expressed in terms of a two-dimensional coordinate system that covers the shell), which happens to be very small, the TST deformation free energy  $\Delta F$  is a simple functional of  $\zeta(\mathbf{r})$  in the form of an integral over the shell surface:

$$(23) \quad \Delta F = \int dS \left\{ \frac{1}{2} \kappa (\Delta \zeta)^2 + \frac{1}{2} \tau (\nabla \zeta)^2 + \frac{1}{2} Y \left( \frac{2\zeta}{R} \right)^2 \right\},$$

where the first term describes the bending-energy cost of the indentation, the second term represents the work by the probe against the genome osmotic pressure  $\Pi$  with  $\tau = \frac{\Pi R}{2}$ , an effective surface tension. The third term measures the stretching of the layer induced by the force with the two-dimensional Young modulus  $Y$  of the layer. To obtain the FDC, the functional derivative of the deformation free energy ( $\frac{\delta \Delta F}{\delta \zeta(\vec{r})}$ ), which results in the differential equation:

$$(24) \quad \frac{\delta \Delta F}{\delta \zeta(\vec{r})} = f(\vec{r}),$$

where  $f(\mathbf{r})$  is the radial force per unit area, exerted by the probe. (24) can be solved analytically in the case of a point force, i.e.  $f(\vec{r}) = F \delta(\vec{r})$ , which creates a linear FDC and a dimple with the radius of order  $\sqrt{R l_b}$  ( $l_b$  being the length scale,  $l_b = \sqrt{\kappa Y}$ ). That is consistent with experiments. For weak applied forces, the shell behaves like a harmonic spring and the measured FDC is indeed linear in many cases.

Larger indentations are more tricky, for calculating the FDC of TST in the nonlinear regime requires the solution of a pair of nonlinear differential equations (so called Föppl - von Kármán (FvK) equations), which is a time-consuming process, and instead of that, finite-element modelling (FEM) is used. The conclusions drawn from FEM are that the elastic response of the non-uniform icosahedral shells might differ from that of uniform spherical shells.

But, mechanical properties of full viruses, with genome enclosed in the protein shell, are much more interesting than the mechanical properties of empty shells. The density of the close-packed genome material inside the water-permeable shells can be so high that it generates significant osmotic pressures  $\Pi$ , in the range of tens of atmospheres. This pressure generates a non-specific tension  $\tau$  along the shell according to Laplace's law,  $\pi = \frac{2\tau}{R}$  (for a spherical shell), which increases the shell's spring constant. The impact of osmotic pressure on the non-specific shell stiffening was investigated for phage  $\lambda$ . Comparing the mechanical properties of empty and full particles with mutant particles

that had a shorter genome (78 and 94% of the wildtype genome), a conclusion was reached - the presence of the dsDNA in  $\lambda$  was noticeable only at very high genome densities. The same experiments that had HSV1 as the subject (a dsDNA virus with structural similarities to tailed dsDNA phages), had odd observations - namely, that full and empty HSV1 capsids showed no mechanical difference in the stiffness measurement. That was probably because the increased stiffness due to the DNA-induced osmotic pressure was too small when compared with the intrinsic stiffness of the capsid shell. However, there are viruses for which the genome-induced shell stiffening effects are very pronounced.

## LITERATURE

Hagan MF (2013) *Modeling viral capsid assembly*. Advances in Chemical Physics; 155:

Katen S, Zlotnick A (2009) *The Thermodynamics of Virus Capsid Assembly*. Methods Enzymol.; 455: 395–417.

Roos WH, Bruinsma R, Wuite GJL (2010) *Physical virology*. Nature Physics; 6: 733-743

Šiber A, Božič AL, Podgornik R (2012) *Energies and pressures in viruses: contribution of nonspecific electrostatic interactions*. Phys.Chem.Chem.Phys. 14, 3746

Zandi R, van der Schoot P, Reguera D, Kegel W, Reiss H (2006) *Classical Nucleation Theory of Virus Capsids*. Biophysical Journal Vol. 90, Issue 6, pp. 1939-1948

## PICTURES

1: [http://bio1151b.nicerweb.net/Locked/media/ch18/18\\_04virus.jpg](http://bio1151b.nicerweb.net/Locked/media/ch18/18_04virus.jpg)

2: Hagan MF (2013) *Modeling viral capsid assembly*. Advances in Chemical Physics, 155;

3: Left: Donadini R, Liew CW, Kwan AH, Mackay JP, Fields BA (2004). *Crystal and solution structures of a superantigen from Yersinia pseudotuberculosis reveal a jelly-roll fold*. Structure 12(1):145-56.

Middle: Krupovic M, Bamford DH (2011) *Protein Conservation in Virus Evolution*. In: eLS. John Wiley & Sons Ltd, Chichester.

Right: Vanlandschoot P, Cao T, Leroux-Roels G (2003) *The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure*. Antiviral Research, Volume 60, Issue 2, Pages 67-74

4: Vanlandschoot P, Cao T, Leroux-Roels G (2003) *The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure*. Antiviral Research, Volume 60, Issue 2, Pages 67-74

5: Hagan MF (2013) *Modeling viral capsid assembly*. Advances in Chemical Physics, 155;

6: Hagan MF (2013) *Modeling viral capsid assembly*. Advances in Chemical Physics, 155;

7: Šiber A, Božič AL, Podgornik R (2012) *Energies and pressures in viruses: contribution of nonspecific electrostatic interactions*. Phys.Chem.Chem.Phys. 14, 3746